Research Paper ■

Effects of a Decision Support System on the Diagnostic Accuracy of Users: A Preliminary Report

ARTHUR S. ELSTEIN, PHD, CHARLES P. FRIEDMAN, PHD, FREDRIC M. WOLF, PHD, GWEN MURPHY, PHD, JUDITH MILLER, MS, PAUL FINE, MD, PAUL HECKERLING, MD, TOM MILLER, MD, JAMES SISSON, MD, SEMA BARLAS, PHD, KEVIN BIOLSI, PHD, MACY NG, MA, XIAO MEI, MA, TIM FRANZ, MA, AMY CAPITANO, MA

Abstract **Objectives:** To assess the effects of incomplete data upon the output of a computerized diagnostic decision support system (DSS), to assess the effects of using the system upon the diagnostic opinions of users, and to explore if these effects vary as a function of clinical experience.

Design: Experimental pilot study. Four clusters of nine cases each were constructed and equated for case difficulty. Definitive findings were omitted from the case abstracts. Subjects were randomly assigned to one of four clusters and were trained on the DSS prior to use.

Subjects: The study involved 16 physicians at three levels of clinical experience (six general internists, four residents in internal medicine, and six fourth-year medical students), from three academic medical centers.

Procedure: Each subject worked up nine cases, first without and then with ILIAD consultation. They were asked to offer up to six potential diagnoses and to list up to three steps that should be the next items in the diagnostic workup. Effects of DSS consultation were measured by changes in the position of the correct diagnosis in the lists of differential diagnoses, pre- and post-consultation.

Results: The DSS lists of diagnostic possibilities contained the correct diagnosis in 38% of cases, about midway between the levels of accuracy of residents and attending general internists. In over 70% of cases, the DSS output had no effect on the position of the correct diagnosis in the subjects' lists. The system's diagnostic accuracy was unaffected by the clinical experience of the users.

■ **JAMIA.** 1996;3:422–428.

To what extent does a diagnostic decision support system (DSS) change or alter the diagnostic reasoning of clinicians? How much do varying levels of clinical expertise affect the output of their interaction with a system? How do clinicians interpret the output? Used by typical clinicians, do these systems offer useful advice? This paper reports a pilot study, the first part of a larger study investigating these and other questions.

Affiliations of the authors: Department of Medical Education, University of Illinois at Chicago (ASE, SB, KB, MN, XM, TF); Laboratory for Computing and Cognition, University of North Carolina (CPF, GM); Laboratory for Computing, Cognition and Clinical Skills, University of Michigan (FMW, JM, AC); Department of Internal Medicine, University of Michigan (PF, JS); Department of Internal Medicine, University of Illinois at Chicago (PH); Department of Internal Medicine, University of North Carolina (TM).

Supported in part by a grant from the National Library of Medicine, RO1-LM 5630.

Correspondence and reprint requests: Arthur S. Elstein, PhD, Department of Medical Education, (m/c 591), University of Illinois at Chicago, 808 S. Wood Street, Chicago, IL 60612.
e-mail: aelstein@uic.edu

Received for publication: 5/8/96; accepted for publication: 7/19/96.

Background

Automated decision supports can perform a variety of functions. These systems range from computer-based alerts and reminders, through computer-based protocols and guidelines, to knowledge-based systems confined to fairly narrow domains, and on to systems intended to provide advice across a very large domain of clinical problems. In our research, we are concerned with the subset that are intended to provide diagnostic aid across a very broad domain, often the entire spectrum of internal medicine. Examples of such broad-spectrum systems are DXplain, Meditel, ILIAD, and QMR.[1]

In the early days of the development of such DSSs, they were commonly referred to as "artificial intelligence" or "expert" systems. The concept underlying these terms seems to have been an "oracle" model of automated expertise: relevant facts or clinical findings were the input, and the output was a proposed solution or set of solutions to a diagnostic problem.[2,3] Clinicians often did not interact directly with the system. A computerized consultation was accomplished by filtering the clinical findings through an expert user of the program, not necessarily a physician. The problems most often studied were diagnostically difficult cases, such as clinicopathological conferences (CPCs) from the New England Journal of Medicine.[4] The criterion for the success of the automated inference system in a number of evaluation studies has been whether the correct diagnosis was relatively high (top 1, top 10, or top 15) on the list of diagnoses proposed by the program, given the input.

By 1989, there was a shift away from this oracular conception of automated clinical inference toward an appreciation that these computer programs had more limited capabilities. This recognition led to a new research emphasis on how clinicians interact with a DSS and what use they make of the advice or suggested diagnoses provided. The term "decision support system" was introduced to replace "expert system" or "artificial intelligence" because it more accurately reflected what could be expected: the reasoning of the human clinician would be supported and, it was hoped, enhanced by the automated expert, but it was unlikely that the role of the diagnostician would be taken over by computer programs.

There seem to be at least three reasons for the shift in emphasis. First, it gradually became clear that the knowledge base of even the most comprehensive expert system was incomplete, just like the knowledge base of a human clinician. Given these inevitable limitations in knowledge and possibly in inference rules and strategies, it was unreasonable to expect the program to behave like an infallible oracle.[3] Second, because of limitations of time and the program's vocabulary, it was often impossible for a physician to convey a complex understanding of a case to a computer program. The physician simply could not tell the program what a human consultant could be told in natural language. In a personal encounter with a contemporary DSS, Kassirer[5] found this feature to be of particular concern. Third, in realistic clinical practice, definitive or highly diagnostic clinical data might not be available for input to a decision support tool. Unless a clinician were considering a particular diagnosis, special diagnostic tests that were needed to confirm or rule out that diagnosis might not be ordered. Thus, the issue is not how well the DSS reasons to a conclusion from a complete data base. Rather, given the necessarily incomplete data base that a puzzled clinician might have assembled in the workup of a diagnostically challenging case, to what extent does the DSS improve the quality of the differential diagnosis and/or suggest the relevant clinical findings needed to reach a more definitive diagnosis?

In the study of INTERNIST I cited earlier, the input was the entire body of clinical data available to the case discussants excluding the definitive findings presented by the pathologists presenting the final diagnoses. In a more recent evaluation of four decision support systems,[1,6] Berner and her colleagues attempted to include all the data provided in written case descriptions, not only pertinent findings. These case descriptions, written by nationally recognized consultants who were not part of the DSS development team, omitted data collected at the consultant's direction. These omissions usually included the definitive tests that confirmed the diagnosis. (In this respect, Berner's studies resemble ours.) Despite considerable effort by the developers of the four programs to translate the vocabulary of the case descriptions into the language of the program, some data could only be approximated in some programs, and some findings could not be entered at all.

Bankowitz and colleagues[7] studied the question of user variability in entering findings into a DSS. Using Quick Medical Reference (QMR) as the index system, they compared the data entry of six physicians with that of the primary developer. They found fairly good agreement in entering positive findings, but less satisfactory agreement on entering negative findings. They did not study the effects of this variation on the output of the DSS.

The QMR investigators have also studied the effects of decision support consultation upon clinicians' diagnoses and management of cases.[8-10] In these studies, however, the QMR consultation was provided by

an expert user of the system, a member of the QMR team, and the physicians caring for the patients did not interact directly with the system. Johnston and her colleagues[11] published a critical appraisal of the literature on clinical decision support systems. They noted that, in most of the studies reviewed, clinicians did not use the computers themselves but were given printed reports generated by an expert-user interaction with the DSS. Five studies of computer-aided diagnosis met their inclusion criteria; four showed a positive effect on correct diagnosis or referral.

The investigation reported here is part of a larger study designed to examine the effect of obtaining a consultation with a broad-based DSS upon the reasoning of physicians with three different levels of clinical experience. The evaluation strategy is designed to be applicable to any DSS of this type. The system examined in this study is ILIAD,[12] developed at the University of Utah. Using cases with diagnoses that are in the knowledge base of the DSS selected, we ask the following questions: What effect does inputting incomplete data have on the diagnostic advice the system offers? What is the impact of the DSS consultation on the diagnostic opinions of users who are physicians but are not expert in using the DSS? Does the impact vary with the user's level of clinical experience? Subsequent research will investigate these questions with QMR.

The entire study is being conducted at three sites: the University of Illinois Health Sciences Center in Chicago, the University of Michigan Medical Center in Ann Arbor, and the University of North Carolina Medical Center in Chapel Hill. All three sites are major academic medical centers, with residency programs in a wide range of specialties. All offer educational programs in nursing, public health, and many associated health professions, such as physical therapy, occupational therapy, etc. All encounter a mixture of patients; some patients are referred for tertiary care, while others use the facility for primary and secondary care. The investigative group at each site includes a research psychologist experienced in medical informatics, a clinician co-investigator, and collaborators responsible for day-to-day operations and data collection. Communication between sites has been maintained by regularly scheduled conference calls, e-mail, and periodic face-to-face meetings.

Methods

Case Materials

At each site, twelve diagnostically challenging cases were selected from recent admissions to the internal medicine service. All cases had discharge diagnoses listed in ILIAD's knowledge base. Cases were first rated by the clinician at the site of origin on a seven-point scale of difficulty, and cases with the lowest two ratings were not considered any further. Only cases rated from three to seven for level of difficulty were retained for the study.

When a case was considered for inclusion in the study, the clinical investigator at each site (PF, PH, TM) wrote a 2- to 3-page case abstract that included all salient history and physical findings (and some not so salient), as recorded in the patient's chart, and all laboratory tests except those the abstracter considered to be gold standard or definitive laboratory tests or diagnostic studies (such as an elevated leukocyte count for a case with right lower quadrant tenderness and a discharge diagnosis of acute appendicitis). Definitive findings (which might be items in the history or physical examination) were omitted for two reasons:

1. The clinical investigators judged that, given the results of definitive diagnostic studies, the cases would not be diagnostically challenging, certainly not for an experienced clinician and probably not for a DSS. These definitive findings (usually test results, but not necessarily) provide the answer to the diagnostic puzzle.
2. A clinician who has these definitive findings or tests would most likely not seek a DSS consultation. In particular, where special tests are at issue, we reasoned that a DSS consultation or analysis of the case would be needed only when the test results are unavailable, or when it is not clear what tests should be ordered, or when a patient's presentation is so unusual that the physician is very perplexed. Consequently, these findings should not be part of a case abstract in a study designed to assess the value added by a DSS.

The case summaries from each site were then distributed to the clinical investigators at the other sites, who independently rated them for level of difficulty. No formal criteria were used to make these judgments, since our purpose was not to assess inter-judge reliability. Instead, our purpose was simply to exclude very easy or trivial diagnostic problems from the spectrum of cases to be included in the study. The cases judged acceptable by the clinician at the site of origin were independently rated for difficulty by the other two clinicians. The generalizability coefficient, Cronbach's alpha,[13] on the initial ratings was 0.63, indicating a moderate level of inter-judge consistency. We found that the judges differed by three or four points (on the seven-point scale) on eight cases. Pre-

dictably, a conference call held to discuss these cases led to some narrowing of the differences. After this call, all final judgments were no more than two points apart, and the generalizability coefficient rose to 0.83, indicating a very satisfactory level of agreement. The mean of their three final ratings was then used as the difficulty rating for each case. These ratings ranged from 3.17 to 6.33. Upon reviewing these ratings, we found that, overall, the most diagnostically challenging cases were either an atypical presentation of a disease or a case with so many diseases or problems that it was difficult to determine whether the presenting complaint was a manifestation of a new disease or a complication of pre-existing conditions.

To design a task of manageable length for each subject, the 36 cases were divided into four clusters of nine cases each. To equate the clusters for difficulty, the 12 cases from each site were divided into three levels of difficulty (low, intermediate, high) with four cases at each level. Each cluster was given three cases from each site: one case of low, one of intermediate, and one of high difficulty. The mean difficulty rating of clusters ranged from 4.63 to 4.72, so by this measure they were effectively equated for case difficulty.

DSS

A frame-based system, ILIAD[12] utilizes both Boolean and Bayesian logic to draw inferences from data. The knowledge base contains over 920 diagnoses and more than 10,000 disease findings.

Subjects

This pilot study involved 16 subjects at three different levels of clinical experience: six general internists, all in sections of general internal medicine in an academic medical center; four residents in internal medicine, all in their second year of postgraduate training; and six fourth-year medical students.

Training

Prior to exposure to the experimental cases, each subject was trained individually to use ILIAD. A standardized training program was developed and used. It provided the subjects with experience in entering data, consulting disease frames, and interpreting the DSS output. Each subject worked up two practice cases as part of the training.

Experimental Procedure

Each subject worked up a cluster of nine cases, first without and then with an ILIAD consultation. Subjects were randomly assigned to a cluster. To avoid the possible effects of order of presentation of cases within a cluster, the order was randomized for each subject, although the first two cases administered were always from the easy level.

For each case within a cluster, each physician initially worked up the case without ILIAD, using the abstracts provided. The participants were instructed to offer a differential diagnosis list of up to six diagnoses, to list up to three steps that should be the next items in the diagnostic workup, and to indicate the likelihood of their seeking a diagnostic consultation for this case (from a human clinician or a DSS) on a four-point scale (4 = definitely, 3 = probably, 2 = possibly, 1 = no). After completing the initial pass through the case, they were given access to ILIAD and could enter whatever findings they chose into the system. They also had an opportunity to use some of the specialized features of the program, such as the critique mode, advice on tests to order, information on the interpretation of positive and negative likelihood ratios, etc. After concluding the DSS consultation, they were again asked to write their differential diagnoses and relevant next steps in the diagnostic workup and to rate the helpfulness of the DSS on a four-point scale (4 = very helpful, 1 = not very helpful). Subjects continued working up cases this way until all nine in the cluster were completed. Sessions varied in length, depending on the time the physician could allocate to the project. Some were as short as one hour, others lasted up to four hours.

In this report, we present evidence that the cases were indeed difficult for the subjects, examine the effects of incomplete data on the DSS output, and explore the effects of the DSS consultation on the participants' differential diagnoses. These preliminary analyses are based on 16 subjects each analyzing nine cases, for a total 144 case workups, first without and then with DSS consultation.

Results

Case Difficulty

Before beginning our analysis of the central research questions, we first sought to confirm that the cases were really diagnostic challenges for our subjects. To answer this question, we examined how often the subjects said they would seek a diagnostic consultation and how often the correct diagnosis was included in their preconsultation lists.

A four-point scale was used to rate the probability of seeking an outside diagnostic consultation. The attending physicians gave ratings of "definitely" and "probably" on 57% of their cases (30/54). The residents gave these ratings on 81% of their cases, and

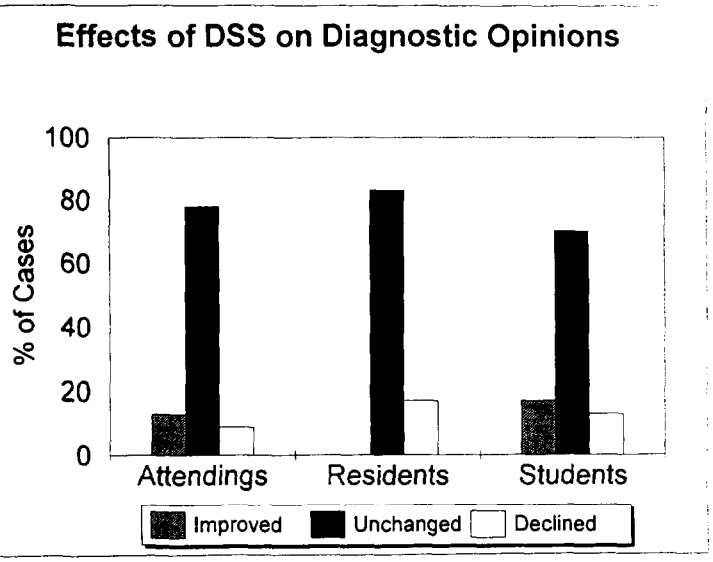

Figure 1 Effects of a decision support system on accuracy of diagnostic opinions.

the medical students gave these ratings on 87%. Thus, the probability of asking for a consultation increases as the amount of clinical experience decreases, but even the most experienced group—general internists on medical school faculties—said they would seek a consultation in more than half of the cases.

Diagnostic accuracy before consultation shows a similar pattern. Diagnostic accuracy was assessed by a very simple measure: the presence of the correct diagnosis anywhere on the subject's list of up to six possible diagnoses. In making these judgments, the discharge diagnosis on the case was adhered to closely. The overall accuracy for attending physicians was 43%; for residents, 33%; and for medical students, 15%. Thus, diagnostic accuracy rises with increasing clinical experience, but no group found the cases easy in the sense of getting 75%–80% correct diagnoses.

Effects of Incomplete Data on the DSS

The first research question was, "What is the effect of incomplete data upon the output of the DSS?" The ILIAD lists of diagnostic possibilities contained the correct diagnosis in 38% of cases (attending physicians and students, 37%; residents, 39%). Since all cases used diagnoses in the ILIAD knowledge base, it follows that ILIAD's inference engine is quite sensitive to the diagnostic impact of the findings that were deliberately omitted.

Effects of the DSS on Diagnostic Opinions

The second research question was, "What is the impact of the DSS consultation upon the diagnostic opinions of the participating physicians?" To answer this question in a preliminary fashion, we examined changes in the position of the correct diagnosis in the lists of differential diagnoses before and after DSS consultation.

The possible effects were classified into three categories: improved, unchanged, or declined. A postconsultation differential diagnosis list was considered "improved" if the rank of the correct diagnosis was higher on the postconsultation list than on the preconsultation list *or* if the correct diagnosis was on the postconsultation list but was not on the preconsultation list. A list was classified as "unchanged" if the rank of the correct diagnosis was unchanged on the two lists or if the correct diagnosis was omitted from both lists. A list was considered to have "declined" if the rank of the correct diagnosis was lower postconsultation than before *or* if the correct diagnosis had been on the preconsultation list but was not on the postconsultation list.

Effects of Level of Clinical Experience

The third research question was, "Does the impact of the DSS vary with the level of clinical experience?" The effect of the DSS on ranking the correct diagnosis is shown in Figure 1. It is clear that in most cases (78% for attending physicians, 83% for residents, 70% for medical students), the DSS had no effect on the position of the correct diagnosis. In about 15% of cases, there was some improvement for faculty physicians and medical students, but no residents' rankings improved. In about 12% of cases overall (9% for attending physicians, 17% for residents, 13% for students), the quality of the lists declined, using the criteria we have defined.

Discussion

This pilot study explored the effects of a DSS (ILIAD) on the diagnostic reasoning of clinicians at three different levels of clinical experience. The cases were intended to be diagnostically challenging, and this goal was achieved. Even the most experienced physicians in the sample reported that they would seek a consultation in more than half of the cases, and the number who indicated a consultation would be sought increased with declining clinical experience.

In this study, the subjects first read case summaries unaided and drew up a differential diagnosis list and a list of relevant next steps. They then entered case findings into the DSS to generate another list of diagnostic possibilities. To study the effects of the computerized DSS upon the diagnostic reasoning of the

clinicians, the change in the position of the correct diagnosis between the two lists was analyzed. Despite the demonstrated difficulty of the cases, the predominant result was that the lists of possible diagnoses generated by the computer consultations had no effect on the position of the correct diagnosis in the subjects' lists.

The ILIAD lists of diagnostic possibilities contained the correct diagnosis in 38% of cases. Using this criterion, ILIAD's diagnostic accuracy was unaffected by the clinical experience of the users. The DSS performed at about the same level of accuracy for all three groups of subjects. Since we have shown that the subjects' differential diagnoses were unchanged in most cases, it follows that ILIAD could have helped in about one third of cases by offering the correct diagnosis but that this advice was either unrecognized or ignored. On the other hand, if the system cannot propose the correct diagnosis in two thirds of cases, in the absence of definitive findings, one can understand why clinicians may have learned or decided to ignore its output on the occasions when it was correct. There is no obvious way for a user to distinguish between a list that contains the correct diagnosis and one that does not.

Based on an admittedly small sample, there is no evidence of differences across the three groups in the quality of the diagnostic advice provided by ILIAD. A clear implication of the data in Figure 1 is that the system's ability to provide correct diagnoses is little affected by clinical experience of the user. Based on our observations of users entering data, we believe that the reason experience has little effect upon the system is this: while experienced users input data more selectively, and fourth-year medical students tend to input everything they can, most of the "excess" data input by the inexperienced clinicians has little diagnostic value. Neither more carefully selected input nor a more scatter-shot approach appears to help the system reach correct diagnoses more often. From one point of view, this finding provides reassurance that one need not worry that the system's accuracy will decline as a result of the user's inability to filter out irrelevant data. On the other hand, the system does not appear to respond with increased accuracy to a more selective, presumably more thoughtful approach to input.

This report has four limitations:

1. **Small sample.** Since this was a pilot study, designed primarily to test the feasibility of methods for data collection and to refine the scoring system, the sample of physicians studied is quite small. While we believe that the sample is representative of the groups from which they were drawn, the conclusions are necessarily tentative. We are presently collecting and analyzing data from a larger sample, using the case clusters and procedures described. Formal analytic statistics could not have been meaningfully used with the small sample of subjects in this pilot. These methods will be applied to the larger data set.
2. **Possible insensitivity of measure of diagnostic accuracy.** To measure the effect of the DSS upon the diagnostic opinions of users, this analysis focused entirely on the position of the correct diagnosis in the differential diagnosis list, and a fairly strict criterion for a match between the "correct diagnosis" (as defined by the discharge diagnosis) and the user's responses was used. This strategy may understate the diagnostic accuracy of both the diagnostic DSS and the human clinicians. For one thing, we may have undercounted nearly synonymous terms. Further, other diagnostic possibilities—close cousins of the correct diagnosis—might have been suggested by the DSS, or their position on the differential diagnosis might have been altered as a result of the consultation. A more sensitive measure of diagnostic accuracy would attend to both the quality of diagnostic hypotheses—how close they are conceptually to the correct diagnosis—and their position in the subjects' personal lists. Thus, the overall accuracy of a differential diagnosis can be conceptualized as a function of the quality and location of the component diagnoses. An analysis of whether this measure of overall accuracy of the users' differential diagnosis is affected by the DSS is now under way.
3. **Relevant next steps.** In the context of an incomplete workup, lacking crucial items from a history or physical examination or definitive laboratory and diagnostic studies, the usefulness of a DSS may lie as much in suggesting relevant missing data as in proposing diagnostic hypotheses. Like other decision support systems, ILIAD provides lists of diagnostic hypotheses by default. Suggested next steps are available, but an extra step is needed to display these; they are not provided by default. These suggestions may be valuable in altering the user clinician's diagnostic plan. For example, if a DSS prompts a clinician to order a definitive diagnostic test, this is in effect almost equivalent to suggesting the correct diagnosis. The effect of the DSS upon the plan of the workup was not examined in this pilot study, but it is part of our program for the full study now under way.
4. **The subjects were all novices in using the DSS.** We do not know if experienced DSS users would

have had different results with these cases. However, since clinicians are typically not experienced users of such systems, we believe results from the subjects we have selected will generalize better to clinicians at comparable levels of clinical experience.

Conclusions

Under the conditions provided in this study—diagnostically challenging cases and incomplete data—the ability of this DSS to reach the correct diagnosis is limited. It performs approximately midway between the residents' and the attending general internists' levels of accuracy. The rate observed is about the same as that reported by Berner et al.[4] for ILIAD and three other DSSs, when gold standard tests were omitted.

In over 70% of cases, the DSS output had no effect on the subjects' diagnostic accuracy. In about 35% of the cases in which the DSS had no effect on the clinicians' reasoning, the correct diagnosis was displayed but was unrecognized or ignored. We do not yet know if a DSS can offer useful advice about the best diagnostic tests to order, given the incomplete data provided.

Data are now being collected from a larger sample of clinicians at all three levels of experience. An improved method of scoring diagnostic accuracy, incorporating both quality and ranking of diagnostic hypotheses, has been developed and will be applied to these data. The effect of DSS consultation on plans for next steps in the workup will also be analyzed.

We acknowledge the assistance of David Potts and Keith Cogdill in preparing cases and running subjects. Earlier versions of this paper were presented at a Workshop on the Evaluation of Knowledge-Based Systems held at the National Library of Medicine, Bethesda, Maryland, December 7–8, 1995, and at the biennial meeting of the European Society for Medical Decision Making, Torino, Italy, June 17–18, 1996.

References ■

1. Berner ES, Webster GD, Shugerman AA, et al. Performance of four computer-based diagnostic systems. N Engl J Med 1994;330:1792–6.
2. Miller RA, Masarie FE. The demise of the "Greek Oracle" model for medical diagnostic systems. Methods Inf Med. 1990;29:1–2.
3. Miller RA. Medical diagnostic decision support systems—past, present and future: A threaded bibliography and brief commentary. J Am Med Inform Assoc. 1994;1:8–27.
4. Miller RA, Pople HE, Myers JD. INTERNIST-I, an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med. 1982;307:468–76.
5. Kassirer JP. A report card on computer-assisted diagnosis—the grade: C. N Engl J Med. 1994;330:1824–5.
6. Berner ES, Jackson JR, Algina J. Relationships among performance scores of four diagnostic decision support systems. J Am Med Inform Assoc. 1996;3:208–15.
7. Bankowitz RA, Blumenfeld BH, Giuse-Bettinsoli N, et al. User variability in abstracting and entering printed case histories with Quick Medical Reference (QMR). In: Stead WW (ed). SCAMC Proc. 1987;68–73.
8. Bankowitz RA, McNeil MA, Challinor SM. A computer-assisted medical diagnostic consultation service: implementation and evaluation of a prototype. Ann Intern Med. 1989; 110:824–32.
9. Bankowitz RA, McNeil MA, Challinor SM, Miller RA. The effect of a computer-assisted general medicine diagnostic consultation service on housestaff diagnostic strategy. Methods Inf Med 1989;28:352–6.
10. Bankowitz RA, Lave JR, McNeil MA. A method for assessing the impact of a computer-based decision support system on health care outcomes. Methods Inf Med. 1992;31:3–11.
11. Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome: A critical appraisal of research. Ann Intern Med. 1994;120:135–42.
12. Applied Medical Informatics. ILIAD 4.2 User Manual. Salt Lake City, UT: Applied Medical Informatics, 1994.
13. Cronbach LJ. Essentials of psychological testing, 3rd ed. New York: Harper & Row, 1970.